PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A method-comparison study regarding the validity and reliability of the Lactate Plus hand-held lactate meter
AUTHORS	Hunt, Brian; Drevets, Sarah; Drevets, Kathryn; Alford, Micah; Salacinski, Amanda

VERSION 1 - REVIEW

REVIEWER	Ricardo J. Fernandes Auxiliary Professor Centre of Research, Education, Innovation and Intervention in Sport , Faculty of Sport and Porto Biomechanics Laboratory, University of Porto, Porto, Portugal I have no conflict of interest regarding the current study
REVIEW RETURNED	03-Nov-2012

GENERAL COMMENTS	The current manuscript aimed to study the validity and reliability of the Lactate Plus lactate analyser, which could be interesting to the BMJ Open readers and researchers worldwide. My main concern is the originality of the topic once this analyser was validated before. Moreover, there are other topics, described under these lines, that authors should take into consideration. Title It could be briefer. Key messages . “The Lactate Plus portable lactate meter provides valid and reliable measurements of blood lactate concentration”. Is this new? Tanner et al (2010) did not report it before? Abstract . “Objectives: The aims of this study were to: 1) determine the validity and reliability of the Nova Biomedical Lactate Plus hand-held lactate meter”. My previous comment also applies here. . As the number of words was not exceeded, some details should be given: (i) which type of exercise was implemented; (ii) was the test continuous or intermittent? (iii) which methodology was used for assessing lactate threshold? . The values of blood lactate concentration should be given in mM per liter. . “The Lactate Plus analyser provides accurate and reproducible measurements … that can be used to estimate workloads corresponding to blood lactate transitions or absolute lactate concentrations”. And what about exercise intensities under lactate threshold? Could this analyser also be used for light-moderate exercise prescription? Introduction . “…has also been proposed as a measure of metabolic acidosis during fetal examinations”. Is this relevant for the current study?

	Were these examinations carried on with portable hand-held lactate meters? Authors should consider removing this example. . Please provide the range values for sample of blood for bench top analysers as done for portable hand-held lactate meters. . "…difference between the reference and hand-held analyzer can be as much as 1.0 mM" and "represent nearly 10% of the full range of values in some Populations". A reference is welcome here. . "This level of disagreement could be explained by the presence of systematic error, which has gone unexamined in previous studies". As Baldari et al (ref #2) has examined systematic error, authors are advised to rewrite this sentence. In addition, authors also refer two studies (refs #9 and 10) that studied this topic. . "Hand-held meters, …, are designed to sample blood directly from a finger". This idea is repeated through the manuscript. However, blood collection from the ear lobe is also very common. Please re-phrase. . "using a finger stick to draw blood it is not uncommon to require "milking" of the finger to get an adequate sample". Was this described before or is from authors' personal experience? This problem can be solved using a vasodilator cream. . "Given that duplicate samples are standard practice". Was this described before or is from authors' personal experience? . After the specific aims of the study, some hypotheses are welcome. Methods . Lactate analysers are, as referred by the authors, a very important instrument to help in training control and prescription of endurance athletes. Nevertheless, the subjects used in the current study do not seem representative of the high trained athletes. This fact could lower the overall quality of the paper. . Units should be abbreviated as proposed in SI Units (eg. min and s). . The portable lactate meter used in the current study was designated in three different forms along the text: Nova Biomedical Lactate Plus, Lactate Plus and Lactate Plus (Nova Biomedical). Please be consistent. . "As per the manufacturer instructions we used a low…". Please rewrite. . "For the first …YSI 2300". This section is hard to follow. Please rewrite. . Please provide the treadmill reference. . I wonder why it was used a discontinuous graded exercise test since the continuous one is the most proper for assessing physiological parameters (e.g. oxygen consumption, heart rate and blood lactate concentrations) and it is not necessary to stop the exercise to collect blood when performing on a treadmill. Please justify your choice. . Please explain it were not used fixed protocol increments. Was this protocol previously described in the literature? . "…Bland-Altman plots were constructed to allow the reader to …". Authors choose their statistical procedures based on scientific principles or in the readers opinion? Please rewrite. . The 1st paragraph of the Data Analysis section is too descriptive. In our opinion, it should be briefer and some references should be added. . As it is well described that after lactate threshold intensity the blood lactate concentrations assumes an exponential increase, we wonder if the use of 2 linear regressions in the best way to assess lactate threshold. If authors want to go deep in this topic, they can consult a study of our group (Fernandes RJ et al.

	Individual Anaerobic Threshold in Swimming, Int J Sports Med 2011; 32: 940–946). . The use of fixed blood lactate concentrations of 2.5 and 4.0 mM/l should be justified. Why not 3.5 mM/l, as proposed by Heck et al (Int J Sports Med 1985; 6: 117-130) for lactate threshold, or 8.0 mM/l that is considered a good indicator of aerobic power? Results . Fig 1: if this is an example of a subject please clearly state it. Moreover, if is important to check if the number of points for the YSE and Lactate Plus are correct (6 and 8, respectively). . It is stated that from the 242 blood samples taken using the hand-held analyser, 27 resulted in error messages due to insufficient sample. This is odd once some portable analysers emit an auditory signal when the quantity of blood is sufficient. Comment, please. Discussion . “However, differences of almost 1.0mM can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy”. This topic should be better developed since it is important to evidence why differences of ~1.0mM/l are so important for training characterization. . Although not being the main focus of the current research, it seems important to give the mean (SD) values for blood lactate concentrations corresponding to lactate threshold. As referred in the text, this parameter is of fundamental importance for endurance athletes; so, it should be presented (and discussed). References: . Please consider to include some relevant studies in accordance with the previous comments.

REVIEWER	Cosme Franklim Buzzachera, Ph.D. Associate Professor, North University of Parana Londrina, Brazil I declare I has no competing interests to disclose.
REVIEW RETURNED	17-Nov-2012

THE STUDY	Minor concern(s): The following minor concerns are presented in order of appearance. Page5, Lines 34: It should be noted the study sample must be enough to validate any instrument. So what about sample size? I believe a sample size calculation should be included in the Methods section. Page7, Lines 43: There is concern with the procedures used to identify the lactate threshold of the participants. In particular, the authors have stated “the threshold was estimated by plotting [lactate] against GXT state. These graphs were visually inspected to determine the lines of best fit”. However, other procedures should be conducted to correctly identify lactate threshold. For example, the visual interpretation of each graph should be independently (and preferentially) made by at least two trained researchers to locate “the point at which blood [lactate] began to increase in a nonlinear fashion” (Beaver’ method, J Appl Physiol, 1985). If the independent determinations of the stage at lactate threshold differ between researchers, a third researcher should adjudicate the difference by independently determining lactate threshold. The three researchers then jointly should agree on the lactate threshold point. If no agreement about the lactate threshold point occurs, data should be

	rejected (Gaskill et al., Med Sci Sports Exerc, 2001). The authors are encouraged to clearly explain how the visual inspection of the graphs to identification of the lactate threshold was carried out in the investigation under review. If no procedures as previously cited were conducted, I believe that, at a minimum, this problem should be acknowledged as a limitation of this study. Strength and Limitations Section: The sentence "We did not compare either instrument to known lactate standards. This may limit our ability to precisely quantify the accuracy of the hand-held analyzer…" should be included within the Discussion section. I believe this suggestion could be useful for allowing a better comprehension of this limitation by reviewers and future readers.
GENERAL COMMENTS	Summary of investigation: The investigation under review attempted to determine the validity and reliability of the Nova Biomedical Lactate Plus hand-held lactate meter and quantify any fixed and proportional bias. The investigation under review was also designed to determine the effect of any bias on the identification of the lactate threshold, as well as to determine the effect that blood sampling methods have on validity and reliability. A group of fifteen physically active, young men and women (no information about sample size calculation) performed a discontinuous graded exercise test (using no previously established protocol) to volitional exhaustion on a motorized treadmill. Blood samples were taken via finger prick and collected in micro capillary tubes for analysis by the reference instrument (YSI 2300) at the end of each 2-min period stage of the exercise testing. Duplicate samples for the hand-held analyzer were either taken directly from the finger or from the micro capillary tubes. Ordinary least products regression analysis was used to assess validity, reliability, and bias in the Nova Biomedical Lactate Plus hand-held lactate meter. The authors noted the estimates of lactate measurements from both YSI 2300 and Lactate Plus were significantly correlated. The authors also noted the differences between instruments had large variability when blood was sampled from finger, but this variability was enormously reduced when instruments measured blood collected in the capillary tubes. No difference in estimates of the lactate threshold (using the Beaver' method) between instruments was found. Reliability for the Lactate Plus was strong with no proportional bias and small fixed bias. Lactate values during exercise ranged from 1.2 to 16.4mM. The authors suggested that the Nova Biomedical Lactate Plus hand-held lactate meter provides accurate and reproducible measurements of blood lactate concentration. General Comments: The authors are to be commended for a well-written manuscript. The arguments for the study are certainly interesting and timely and greatly advance our understanding of the use of a new hand-held lactate meter in laboratory and/or field settings. However, I believe there are some minor concerns and issues with the manuscript in its current form that need to be addressed before being considered for publication. These are outlined below. It is the hope that the following critique will be received in the manner in which it is delivered and be used to help improve the quality of the manuscript. Minor concern(s): The following minor concerns are presented in order of appearance. Page5, Lines 34: It should be noted the study sample must be enough to validate any instrument. So what about sample size? I

	believe a sample size calculation should be included in the Methods section. Page7, Lines 43: There is concern with the procedures used to identify the lactate threshold of the participants. In particular, the authors have stated "the threshold was estimated by plotting [lactate] against GXT state. These graphs were visually inspected to determine the lines of best fit". However, other procedures should be conducted to correctly identify lactate threshold. For example, the visual interpretation of each graph should be independently (and preferentially) made by at least two trained researchers to locate "the point at which blood [lactate] began to increase in a nonlinear fashion" (Beaver' method, J Appl Physiol, 1985). If the independent determinations of the stage at lactate threshold differ between researchers, a third researcher should adjudicate the difference by independently determining lactate threshold. The three researchers then jointly should agree on the lactate threshold point. If no agreement about the lactate threshold point occurs, data should be rejected (Gaskill et al., Med Sci Sports Exerc, 2001). The authors are encouraged to clearly explain how the visual inspection of the graphs to identification of the lactate threshold was carried out in the investigation under review. If no procedures as previously cited were conducted, I believe that, at a minimum, this problem should be acknowledged as a limitation of this study. Strength and Limitations Section: The sentence "We did not compare either instrument to known lactate standards. This may limit our ability to precisely quantify the accuracy of the hand-held analyzer…" should be included within the Discussion section. I believe this suggestion could be useful for allowing a better comprehension of this limitation by reviewers and future readers.

VERSION 1 – AUTHOR RESPONSE

Responses to Dr. Fernandes

1. **Title could be briefer.**
   We appreciate Dr. Fernandes' sentiments, and would like to have a more concise title as well. However, we have not been able to devise a title of less than 17 words that is both adequately descriptive of the study and meets the journal's requirement to include the study design in the title. We hope that Dr. Fernades will note that our title falls well short of the 50-word limit set by the publishers of BMJOpen.

2. **"The Lactate Plus portable lactate meter provides valid and reliable measurements of blood lactate concentration". Is this new? Tanner et al (2010) did not report it before?**
   Tanner did conclude that the lactate Plus meter "displayed good reliability and accuracy…" However, Tanner's conclusions rely on a questionable analytic approach. Moreover, there clearly appears to be systematic measurement error that was not examined. If one looks at their Figure 4 (shown below) it appears as though a proportional bias exists. This is more evident in Tanner's figure 5 (also shown below). Our approach does not suffer from the assumptions inherent in Tanners analytical approach. Our use of least-products regression allows the reader to assess the accuracy and reliability based on three independent parameters: 1) correlation coefficient; 2) the degree of proportional bias; and 3) the degree of fixed bias. While we come to the same conclusion as Tanner et al, our conclusions are based

on a firm analytical approach. Moreover, our approach indicates the meter is 93% more accurate than reported by Tanner et al.

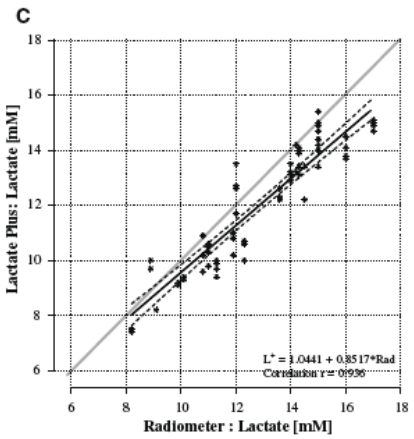

Fig. 4 Portable analyser correlation plots for Lactate Pro (**a**, *circles*), Lactate Scout (**b**, *triangles*) and Lactate Plus (**c**, *asterisks*) analysers versus Radiometer ABL 700 analyser. Linear regression is represented by *solid black line*, ±95% CI by *dashed lines* and line of identity by *grey solid line*. Linear regression equation and correlation coefficients are presented on *bottom right of graph*

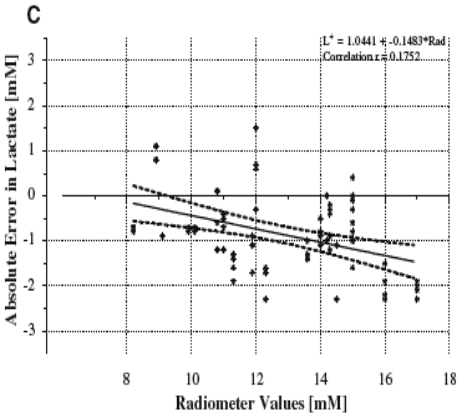

Fig. 5 Portable analyser Bland–Altman plots for Lactate Pro (**a**, *circles*), Lactate Scout (**b**, *triangles*) and Lactate Plus (**c**, *asterisks*) analysers versus Radiometer ABL 700 analyser. Linear regression is represented by *solid black line* and ±95% CI by *dashed lines*. Linear regression equation and correlation coefficients are presented on *top right of graph*

We discuss these points in the Discussion on page 10 as follows: *Though not specifically assessed, it does appear that Tanner's reported difference between the hand held and reference analyzer is significantly influenced by a proportional bias (Ref 8, Figures 4 and 5). The fact that our data shows little proportional bias (Figure 3) may account for the greater agreement between analyzers that we observed. It is possible that if Tanner had been able to independently determine the proportional and fixed bias, their analysis may have revealed a small bias similar to ours. Differences in reference instruments would not likely explain the greater measurement error reported by Tanner, given that their instrument undergoes a 3-point calibration and 2-point calibration check every few hours, similar to our reference instrument.*

Moreover, given that Tanner's figures showed a strong proportional bias, as does most other validity data from various hand-held analyzers, our study took the next obvious step, and tested whether this proportional bias was enough to affect the detection of the lactate threshold.

3. **"Objectives: The aims of this study were to: 1) determine the validity and reliability of the Nova Biomedical Lactate Plus hand-held lactate meter". My previous comment also applies here.**
   See response for comment #2 above.

4. **As the number of words was not exceeded, some details should be given: (i) which type of exercise was implemented; (ii) was the test continuous or intermittent? (iii) which methodology was used for assessing lactate threshold?**
   Your point is well taken. We have added this information to the abstract at follows: ***Design and Participants:*** *In this method comparison study 15 active men and women performed a discontinuous graded exercise test to volitional exhaustion on a motorized treadmill. …* ***Primary Outcome Measurements:*** *… Lactate threshold was determined by visual inspection.*

5. **The values of blood lactate concentration should be given in mM per liter.**
   Thank you for catching this oversight. This has been corrected throughout the manuscript and figures.

6. **"The lactate Plus analyzer provides accurate and reproducible measurements… that can be used to estimate workloads corresponding to blood lactate concentration transitions or absolute lactate concentrations". And what about exercise intensities under lactate threshold? Could this analyzer also be used for light-moderate exercise prescription?**
   As is implicit in our statement quoted above, the lactate measurements from the hand-held analyzer can be used for estimating workloads at *any* absolute lactate concentration. However, we have revised the last sentence in the abstract to make this point more explicit as follows: ***Conclusion:*** *The Lactate Plus analyzer provides accurate and reproducible measurements of blood lactate concentration that can be used to estimate workloads corresponding to blood lactate transitions or any absolute lactate concentrations.*

7. **"…has also been proposed as a measure of metabolic acidosis during fetal examinations". Is this relevant for the current study? Were these examinations carried on with protable hand-held lactate meters? Authors should consider removing this example.**
   We can understand Dr. Fernandes' point, but we included this point for 2 reasons: 1) to help the reader understand that blood lactate measurement is important beyond the narrow field of athletic performance; and 2) to help the reader understand why this paper is appropriate for publication in BMJ. To answer Dr. Fernandes 2nd question, the study cited (Ridenour et al.)

specifically used the Lactate Plus lactate meter to measure blood lactate concentration in order to indicate fetal acidosis.

8. **Please provide the range of values for sample of blood for bench top analyzers as done for portable hand-held lactate meters.**
   Thank you for helping us stay consistent in the development of our thoughts. We have revised the Introduction to as follows: *Portable hand-held lactate meters have advantages over bench top models including: 1) their ability to rapidly sample blood lactate concentration ([lactate]), in or outside the laboratory, 2) they require a much smaller sample of blood (0.5 – 0.7□l) than many bench top analyzers (25 – 75□l), and 3) they can be purchased and operated at a lower cost than many bench top models.*

9. **"…difference between the reference and hand-held analyzer can be as much as 1.0 mM" and "represent nearly 10% of the full range of values in some Populations". A reference is welcome here.**
   This certainly would not be true for highly trained athletes, but can be true for sedentary or untrained individuals. We have included an appropriate reference as follows: *While the majority of studies report the [lactate] measured using hand-held analyzers is similar to those of various bench top models, the mean difference between the reference and hand-held analyzer can be as much as 1.0 mM·l^{-1}. This can represent nearly 10% of the full range of values in some populations.* [9] (Juel C, Klarskov C, Nielsen JJ, Krustrup P, Mohr M, Bangsbo J. Effect of high-intensity intermittent training on lactate and H+ release from human skeletal muscle. *Am J Physiol Endocrinol Metab* 2004;286(2):E245-51.)

10. **"This level of disagreement could be explained by the presence of systematic error, which has gone unexamined in previous studies". As Baldari et al (ref#2) has examined systematic error, authors are advised to rewrite this sentence. In addition, authors also refer two studies (refs #9 and 10) that studied this topic.**
    We respectfully disagree. Perhaps the authors and Dr. Fernandes are using the term "systematic error" differently. We explicitly define our use of the term in the introduction based on the definition of Ludbrook (refs 10 and11). We also clearly describe the biases produced by systematic measurement error, which previous studies have not examined. For example, it is clear to us that the data from Baldari et al. displays a proportional bias as shown in their Figure 2, shown near the top of the next page. Moreover, in Figure 3 Baldari reports regression slopes between 0.938 and 1.105, yet does not report whether these slope are significantly different from 1.0 (indicative of a proportional bias). Thus, Baldari did not look for evidence of systematic measurement error in their data. Furthermore, references 9 and 10 (now refs 10 and 11), to which Dr. Fernandes refers do not examine systematic measurement error in hand-held lactate meters, but are papers describing the advantages of least products regression over the least squares regression approach used by Baldari and most other authors that have performed validation studies on these hand-held meters.

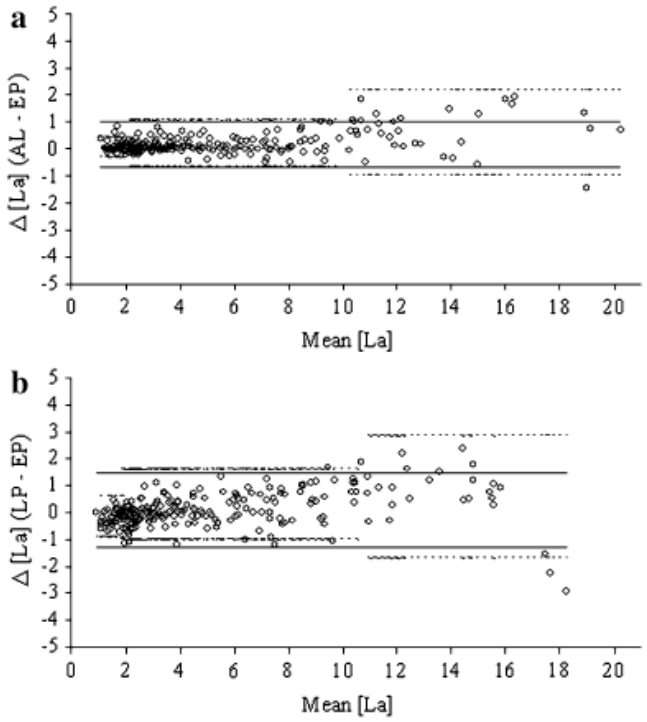

Fig. 2 Bland-Altman plots showing limits of agreement between blood lactate concentrations expressed in mM ([La]) measured using the EBIO plus, Accutrend and Lactate Pro analyzers. *Dashed lines* are the limits of agreement for low, medium, high [La]. **a** Relationship of mean [La] determined by Accutrend and EBIO plus with the difference in La between analyzers (Δ [La] (AL-EP)). **b** Relationship of mean [La] determined by Lactate Pro and EBIO plus with the difference in La between analyzers (Δ [La] (LP-EP))

11. **"Hand-held meters…, are designed to sample blood directly from a finger". This idea is repeated through the manuscript. However, blood collection from the ear lobe is also very common. Please re-phrase.**
    We understand the reviewer's perspective. Therefore, we have revised this statement on page 5 to read as follows: *Hand-held meters, however, are designed to sample blood directly from a puncture for ease of use in the field.*

12. **"using a finger stick to draw blood it is not uncommon to require "milking" of the finger to get an adequate sample". Was this described before or is from authors' personal experience? This problem can be solved using a vasodilator cream.**
The effects of milking of the finger to produce a blood drop for sampling has been most extensively studied with hand-held glucose meters. For example see Fruhstorfer and Quarder. Diabetes Res Clin Pract 85(1), e14-15, 2009. Moreover, the manufacturer of the Lactate Plus meter specifically advises users that if they must squeeze the finger to form a drop of blood "do not squeeze vigorously." We acknowledge that a vasodilating cream could be used to minimize or eliminate the need for milking of the finger, though this is extremely rare in the literature. We have revised the Discussion on page 11 to read: *The milking of the finger to obtain a blood sample can cause the dilution of the blood sample by interstitial fluid. The manufacturer warns the user against vigorous squeezing of the finger to obtain a blood drop. The use of a vasodilating cream may resolve this issue.*

13. **"Given that duplicate samples are standard practice". Was this described before or is from authors' personal experience?**
This is a "best practice" based on statistical principles as well as the relatively large differences reported by investigators such as Baldari (SEE = 0.55 mM.l-1) and Tanner (0.9 mM.l-1). Nonetheless, we have removed the statement from the Introduction on page 5 and the Discussion on page 11.

14. **After the specific aims of the study, some hypotheses are welcome.**
Typically, validity and reliability studies are not hypothesis driven (see Ref. 1 – 8).

15. **Lactate analysers are, as referred by the authors, a very important instrument to help in training control and prescription of endurance athletes. Nevertheless, the subjects used in the current study do not seem representative of the high trained athletes. This fact could lower the overall quality of the paper.**
We disagree that the training status of the study participants has any relevance to this paper. The aim of this study was to assess the accuracy and reliability of the Lactate Plus analyzer. It is unclear why the device would accurately measure lactate concentration in one population and not in another.

16. **Units should be abbreviated as proposed in SI Units (eg min and s).**
Again, thank you for bringing this oversight to our attention. These abbreviations, where they occur, now conform to those proposed for SI units.

17. **The portable lacate meter used in the current study was designated in three different forms along the text: Nova Biomedical Lactate Plus, Lactate Plus and Lactate Plus (Nova Biomedical). Please be consistant.**
We agree that this can be distracting to the reader. We now consistently refer to the lactate meter as Lactate Plus lactate meter.

18. **"As per the manufacturer instructions we used a low…" Please rewrite.**
    We have revised the sentence on page 6 to read as follows: *Following the manufacturer instructions we used a low (1.0 – 1.6 mM·l^{-1}) and high (4.0-5.4 mM·l^{-1}) quality control solution to ensure the lactate meter was operating properly at the beginning of each data collection session.*

19. **"For the first…YSI 2300". This section is hard to follow. Please rewrite.**
    This was indeed a difficult section to write. We appreciate another opportunity to make our writing more clear. We have revised this section on page 6 to read as follows: *For the first nine participants three blood samples were taken directly from the finger between each stage of the graded exercise test (GXT). All samples were taken in this order: 1) hand-held directly from finger, 2) capillary tubes for the YSI 2300 from the finger, and 3) a second sample directly from finger using the hand-held meter. To assess the effect of blood sampling techniques on the accuracy of the hand-held meter blood was drawn from the finger into capillary tubes and allocated to both the YSI 2300 and hand-held meter for the last six participants.*

20. **Please provide treadmill reference.**
    We have provided the reference on page 6 as follows: *Participants performed a discontinuous graded exercise test (GXT) on a motorized treadmill (Quinton TM65).*

21. **I wonder why it was used a discontinuous graded exercise test since the continuous one is the most proper for assessing physiological parameters (e.g. oxygen consumption, heart rate and blood lactate concentrations) and it is not necessary to stop the exercise to collect blood when performing on a treadmill. Please justify your choice.**
    We agree that blood samples can be collected while the subject is walking or running on the treadmill. However, we chose to use a discontinuous protocol because we were not collecting a single blood sample, but three samples. Thus, in our pilot testing we found that a discontinuous protocol allowed us to collect all three samples during the 1-minute sampling period.

22. **Please explain it were not used fixed protocol increments. Was this protocol previously described in the literature?**
    This GXT protocol has not been previously described in the literature. It is unclear to the authors how our protocol would negatively affect our ability to assess the accuracy and reliability of the hand-held analyzer, or model changes in blood lactate concentration.

23. **"…Bland-Altman plots were constructed to allow the reader to …" Authors choose their statistical procedures based on scientific principles or in the readers opinion? Please rewrite.**
We agree that statistical analyses should be chosen based on the experimental question or hypothesis being tested and statistical principles. However, we also believe that an important aspect of writing a scientific paper is to inform the readers. This includes helping the readers understand our findings within the context of previous work. Therefore, we chose to construct a Bland-Altman plot because this has become commonplace in methodological studies. (see ref 1-6,8) So as a service to our readers, we provide a common point of comparison between our data and those previously published.

24. **The 1st paragraph of the Data Analysis section is too descriptive. In our opinion, it should be briefer and some references should be added.**
Based on the analytic approaches used in previous validation studies, it is reasonable to assume that a thorough explanation of our approach is warranted. We have added references to this section on page 7 as follows: *Validity was determined from the correlation coefficient in combination with the presence and degree of bias. The degree of fixed bias was determined from the y-intercept 95% confidence intervals. If the confidence interval for the intercept includes the value of zero, then there is no fixed bias. Proportional bias was determined from the 95% confidence interval for the slope. If the confidence interval for the slope includes the value of 1.0, then there is no proportional bias. Ordinary least products regression gives different slopes and y-intercepts than does least squares regression because error is assumed in both hand-held and bench top analyzers.* [10 11]

25. **As it is well described that after lactate threshold intensity of blood lactate concentrations assumes an exponential increase, we wonder if the use of 2 linear regressions in the best way to assess lactate threshold. If authors want to go deep in this topic, the can consult a study of our group (Fernandes RJ et al. Individual Anaerobic Threshold in Swimming, Int J Sports Med 2011; 32: 940-946).**
This is the one common concern shared by the reviewers, and we agree this is an issue that needs to be addressed. We chose to follow the procedures outlined by Gaskill et al (Med Sci Sports Exerc 2001; 33(11):1841-48) as suggested by Dr. Buzzachera. This has slightly reduced the correlation coefficient and changed the parameters of the regression line. We have clarified our approach in the Methods section on page 7 as follows: *Lactate threshold was defined as the point at which blood [lactate] began to increase in a non-linear fashion.*[12 13] *The threshold was estimated by plotting [lactate] against GXT stage. These graphs were visually inspected to determine the lines of best fit by the two evaluators. The equations for each line were set equal to one another and solved for the point of intersection (Figure 1). The values from each evaluator were averaged.* [14]
We have also revised our results accordingly on page 8 as follows: *there was excellent agreement between estimates of the lactate threshold based on lactate values from the hand-held lactate meter compared to those from the bench top analyzer (r = 0.97). Moreover, there was neither a proportional bias (95% CI for slope: 0.910 to 1.098), nor a fixed bias (95% CI for y-intercept: -0.396 to 0.325) in estimates of the lactate threshold from the hand-held analyzer.*

26. **The use of fixed blood lactate concentrations of 2.5 and 4.0 mM/l should be justified. Why not 3.5 mM/l, as proposed by Heck et al (Int J Sports Med 1985; 6: 117-130) for lactate threshold, or 8.0 mM/l that is considered a good indicator of aerobic power?**
The reviewer's point is well taken. Many investigators use several different absolute lactate values to quantify blood lactate concentration. We have added references to support our use of 2.5 and 4.0 $mM \cdot l^{-1}$ on page xx as follows: *These equations were also used to calculate the stage that corresponded to an absolute blood [lactate] of 2.5 and 4.0 $mM \cdot l^{-1}$.*[14 15]

27. **Fig 1: if this is an example of a subject please clearly state it. Moreover, if is important to check if the number of points for the YSE and Lactate Plus are correct (6 and 8, respectively).**
We have revised the figure legend on page 16 to indicate these data are from a study participant as follows: *Figure 1. Determination of the Lactate threshold by visual inspection. Shown are data from a representative study participant and the lines of best fit that were determined independently for data from the YSI 2300 lactate analyzer and the Lactate Plus lactate meter.*

We were not able to collect any blood after stage 4 and could not get a blood sample with the hand-held analyzer after stage 7. Thus, the YSI data set contains 9 data points and the Lactate Plus data set contains 8 points. Values for rest and stages 1 and 2 are nearly identical and are difficult to distinguish.

28. **It is stated that from the 242 blood samples taken using the hand-held analyzer, 27 resulted in error messages due to insufficient sample. This is odd once some portable analyzers emit an auditory signal when the quantity of blood is sufficient. Comment Please.**
We agree it is odd that the auditory signal can sound and yet still give an error message that is associated with inadequate sample volume. This may be due to operator error, though even when care is taken this still occurs. We have revised the Discussion on page 11 to expand on this point as follows: *We also found that the hand-held analyzer was unable to analyze the blood sample 11% of the time, presumably from an insufficient sample volume. This was surprising given that the Lacate Plus meter provides an audible signal to indicate when the test strip has a sufficient volume of blood for analysis. Our experience has shown that anticipating the filling of the test strip can result in both the audible signal and an error. However, even when great care is taken, one can still get an audible full signal and the error message.*

29. **"However, differences of almost 1.0mM can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy". This topic should be better developed since it is important to evidence why differences of ~1.0mM/l are so important for training characterization.**
We appreciate the point made by Dr. Fernandes. We have developed our point more fully in the Discussion on page 10 as follows: *However, differences of almost 1.0 $mM \cdot l^{-1}$ can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy. Weltman et al. reported that women who trained at an intensity corresponding to about 2.5 $mM \cdot l^{-1}$ showed greater improvement in blood lactate parameters, but less of an improvement in VO2max than did women training at their lactate threshold.* [15] *If true, then an error in the measurement of blood lactate concentration could lead to suboptimal improvements in either lactate parameters or VO2max.*

30. **Although not being the main focus of the current research, it seems important to give the mean (SD) values for blood lactate concentrations corresponding to lactate threshold. As referred in the text, this parameter is of fundamental importance for endurance athletes; so, it should be presented (and discussed).**
The purpose of estimating the lactate threshold was to determine if the proportional bias we anticipated seeing was large enough to affect the estimation of the lactate threshold or other lactate parameters found in the literature. Thus, it seems to us that the mean value and variability of lactate thresholds within our study sample irrelevant to the aims of our study and interpretation of our data. If we were trying to draw some conclusion about the "eliteness" of our study sample it would certainly make sense, but that is not the case here.

31. **Please consider to include some relevant studies in accordance with the previous comments.**
    As can be seen from our responses above, several references have been added to address Dr. Fernandes' concerns.

Responses to Dr. Buzzachera

1. **Page5, Lines 34: It should be noted the study sample must be enough to validate any instrument. So what about sample size? I believe a sample size calculation should be included in the Methods section.**
   Dr. Buzzachera's point is correct; a sample size calculation should have been done a priori. However, a post-hoc sample size calculation is inappropriate. The concern now would be if we reported clinically significant differences, say close to 1 mM.l-1 and reported that our analysis indicated this was not statistically different from zero. This would be indicative of a sample size problem. As can be seen by the results, our sample size was adequate to see a difference of 0.056 $mM \cdot l^{-1}$, a difference that is 93% smaller than had previously been reported. Thus, our sample size seems more than adequate given our statistical approach.

2. **Page7, Lines 43: There is concern with the procedures used to identify the lactate threshold of the participants. In particular, the authors have stated "the threshold was estimated by plotting [lactate] against GXT state. These graphs were visually inspected to determine the lines of best fit". However, other procedures should be conducted to correctly identify lactate threshold. For example, the visual interpretation of each graph should be independently (and preferentially) made by at least two trained researchers to locate "the point at which blood [lactate] began to increase in a nonlinear fashion" (Beaver' method, J Appl Physiol, 1985). If the independent determinations of the state at lactate threshold differ between researchers, a third researcher should adjudicate the difference by independently determining lactate threshold. The three researchers then jointly should agree on the lactate threshold point. If no agreement about the lactate threshold point occurs, data should be rejected (Gaskill at al., Med Sci Sports Exerc, 2001). The authors are encouraged to clearly explain how the visual inspection of the graphs to identification of the lactate threshold was carried out in the investigation under review. If no procedures as previously cited were conducted, I believe that, at a minimum, this problem should be acknowledged as a limitation to this study.**
   Thank you for your comments and guidance. This is the one common concern between reviewers. Please see our response to Dr. Fernandes' comment # 25.

3. **Strength and Limitations Section: The sentence "We did not compare either instrument to known lactate standards. This may limit our ability to precisely quantify the accuracy of the hand-held analyzer…" should be included within the Discussion section. I believe this suggestion could be useful for allowing a better comprehension of this limitation by reviewers and future readers.**
   We understand Dr. Buzzachera's suggestion that this limitation also appear in the discussion, as it should. Thus, we have added this limitation and a further explanation to the Discussion section on page 12 as follows: *We did not compare the Lactate Plus lactate meter to known standards. This limits the precision with which we can quantify the accuracy of the hand-held analyzer. However, our reference instrument was calibrated using 3 known lactate standards across a supraphysiologic range. Our analysis assumes measurement error in both the hand-held and reference instrument. Thus it is likely that by comparing the Lactate Plus lactate meter directly to known lactate standards, our fixed bias would be reduced.*

VERSION 2 – REVIEW

REVIEWER	Ricardo J. Fernandes Centre of Research, Education, Innovation and Intervention in Sport (CIFI2D), Faculty of Sport, University of Porto, Porto, Portugal No conflict of interest
REVIEW RETURNED	28-Dec-2012

THE STUDY	Other references could be included to reinforce the paper.
GENERAL COMMENTS	The authors would like to thank Dr.'s Fernandes and Buzzachera for their insights and suggestions regarding our manuscript. We have carefully considered each comment and it's potential impact on our manuscript. We have responded to each comment below, providing the details of our changes to the manuscript, where each change can be found within the manuscript, or our rationale as to why we have not revised the manuscript as suggested by the reviewer. We believe that the reviewer's comments have helped us write an improved manuscript. Authors have done a good job in following the reviewers' suggestions. The manuscript is stronger now. Nevertheless, there are some points that still should be addressed. Please observe the under referred comments. The points that were adequately followed by the authors were deleted to facilitate the reviewing process. Responses to Dr. Fernandes 1. Title could be briefer. We appreciate Dr. Fernandes' sentiments, and would like to have a more concise title as well. However, we have not been able to devise a title of less than 17 words that is both adequately descriptive of the study and meets the journal's requirement to include the study design in the title. We hope that Dr. Fernades will note that our title falls well short of the 50-word limit set by the publishers of BMJOpen. This reviewer is perfectly aware of the BMJOpen guidelines. Please note that the previous suggestions were not based on personal sense or believes but on scientific writing rules. One of those states that when writing a scientific paper, the language should be completely clear and concise. In this case, there are some words in the title that do not give any additional information to the readers, as it is evident that this is a "study" and that the Lactate Plus is a "hand-held lactate meter". 2. "The Lactate Plus portable lactate meter provides valid and reliable measurements of blood lactate concentration". Is this new? Tanner et al (2010) did not report it before? Tanner did conclude that the lactate Plus meter "displayed good reliability and accuracy…" However, Tanner's conclusions rely on a questionable analytic approach. Moreover, there clearly appears to be systematic measurement error that was not examined. If one looks at their Figure 4 (shown below) it appears as though a proportional bias exists. This is more evident in Tanner's figure 5 (also shown below). Our approach does not suffer from the assumptions inherent in Tanners analytical approach. Our use of least-products regression allows the reader to assess the accuracy

	and reliability based on three independent parameters: 1) correlation coefficient; 2) the degree of proportional bias; and 3) the degree of fixed bias. While we come to the same conclusion as Tanner et al, our conclusions are based on a firm analytical approach. Moreover, our approach indicates the meter is 93% more accurate than reported by Tanner et al. We discuss these points in the Discussion on page 10 as follows: *Though not specifically assessed, it does appear that Tanner's reported difference between the hand held and reference analyzer is significantly influenced by a proportional bias (Ref 8,Figures 4 and 5). The fact that our data shows little proportional bias (Figure 3) may account for the greater agreement between analyzers that we observed. It is possible that if Tanner had been able to independently determine the proportional and fixed bias, their analysis may have revealed a small bias similar to ours. Differences in reference instruments would not likely explain the greater measurement error reported by Tanner, given that their instrument undergoes a 3-point calibration and 2-point calibration check every few hours, similar to our reference instrument.* Moreover, given that Tanner's figures showed a strong proportional bias, as does most other validity data from various hand-held analyzers, our study took the next obvious step, and tested whether this proportional bias was enough to affect the detection of the lactate threshold. The inclusion of these sentences in the Discussion section better justifies the pertinence of the study. 10. "This level of disagreement could be explained by the presence of systematic error, which has gone unexamined in previous studies". As Baldari et al (ref#2) has examined systematic error, authors are advised to rewrite this sentence. In addition, authors also refer two studies (refs #9 and 10) that studied this topic. We respectfully disagree. Perhaps the authors and Dr. Fernandes are using the term "systematic error" differently. We explicitly define our use of the term in the introduction based on the definition of Ludbrook (refs 10 and11). We also clearly describe the biases produced by systematic measurement error, which previous studies have not examined. For example, it is clear to us that the data from Baldari et al. displays a proportional bias as shown in their Figure 2, shown near the top of the next page. Moreover, in Figure 3 Baldari reports regression slopes between 0.938 and 1.105, yet does not report whether these slope are significantly different from 1.0 (indicative of a proportional bias). Thus, Baldari did not look for evidence of systematic measurement error in their data. Furthermore, references 9 and 10 (now refs 10 and 11), to which Dr. Fernandes refers do not examine systematic measurement error in hand-held lactate meters, but are papers describing the advantages of least products regression over the least squares regression approach used by Baldari and most other authors that have performed validation studies on these hand-held meters. The question is not this reviewer interpretation of the term "systematic error" but the existing literature on that topic that uses

	precisely that expression. So, aiming for a clearer text, the above explanation should be given (in a synthetic way) in the manuscript (eventually in the Discussion section). 15. Lactate analysers are, as referred by the authors, a very important instrument to help in training control and prescription of endurance athletes. Nevertheless, the subjects used in the current study do not seem representative of the high trained athletes. This fact could lower the overall quality of the paper. We disagree that the training status of the study participants has any relevance to this paper. The aim of this study was to assess the accuracy and reliability of the Lactate Plus analyzer. It is unclear why the device would accurately measure lactate concentration in one population and not in another. Two questions arise here. The 1st is related to the authors´ writing style: if the Introduction starts with reference to the fact that "blood lactate accumulation is a common measure in the physiological assessment of endurance athletes", naturally the readers will maintain this target population in mind, and will find confusing that no athletes were used during the experimental procedures. Reference to "training intensity or efficacy" when using blood lactate concentration values is also presented in the Discussion section. The 2nd question is related to the usefulness of the tests conducted using the Lactate Plus. If it will be mainly used for assessing lactate threshold (as often referred by the authors along the text) it should be for training control and evaluation purposes. So, as the subjects involved in training process are athletes, the population used in the current study is very relevant. This should be assumed as a study limitation. 22. Please explain it were not used fixed protocol increments. Was this protocol previously described in the literature? This GXT protocol has not been previously described in the literature. It is unclear to the authors how our protocol would negatively affect our ability to assess the accuracy and reliability of the hand-held analyzer, or model changes in blood lactate concentration. The assessment of lactate threshold and other physiologic parameters are strongly affected by the used protocol (there are several examples of lactate threshold assessment protocols in the literature). Thus, the GXT protocol should be better described (e.g. why it were not used fixed protocol increments in the first steps) and discussed with similar (already published) approaches. 23. "…Bland-Altman plots were constructed to allow the reader to …" Authors choose their statistical procedures based on scientific principles or in the readers opinion? Please rewrite. We agree that statistical analyses should be chosen based on the experimental question or hypothesis being tested and statistical principles. However, we also believe that an important aspect of writing a scientific paper is to inform the readers. This includes helping the readers understand our findings within the context of previous work. Therefore, we chose to construct a Bland-Altman plot

	because this has become commonplace in methodological studies. (see ref 1-6,8) So as a service to our readers, we provide a common point of comparison between our data and those previously published. Off course it is important to inform the readers. What should be avoided is to refer it (it is implicit). Please rewrite. 24. The 1st paragraph of the Data Analysis section is too descriptive. In our opinion, it should be briefer and some references should be added. Based on the analytic approaches used in previous validation studies, it is reasonable to assume that a thorough explanation of our approach is warranted. We have added references to this section on page 7 as follows: *Validity was determined from the correlation coefficient in combination with the presence and degree of bias. The degree of fixed bias was determined from the y-intercept 95% confidence intervals. If the confidence interval for the intercept includes the value of zero, then there is no fixed bias. Proportional bias was determined from the 95% confidence interval for the slope. If the confidence interval for the slope includes the value of 1.0, then there is no proportional bias. Ordinary least products regression gives different slopes and y-intercepts than does least squares regression because error is assumed in both hand-held and bench top analyzers. 10 11* The writing style should be perfected, per example by merging consecutive sentences: "Validity was determined from the correlation coefficient in combination with the presence and degree of bias (determined from the y-intercept 95% confidence intervals)". This should be done for all the manuscript. 25. As it is well described that after lactate threshold intensity of blood lactate concentrations assumes an exponential increase, we wonder if the use of 2 linear regressions in the best way to assess lactate threshold. If authors want to go deep in this topic, the can consult a study of our group (Fernandes RJ et al. Individual Anaerobic Threshold in Swimming, Int J Sports Med 2011; 32: 940-946). This is the one common concern shared by the reviewers, and we agree this is an issue that needs to be addressed. We chose to follow the procedures outlined by Gaskill et al (Med Sci Sports Exerc 2001; 33(11):1841-48) as suggested by Dr.Buzzachera. This has slightly reduced the correlation coefficient and changed the parameters of the regression line. We have clarified our approach in the Methods section on page 7 as follows: *Lactate threshold was defined as the point at which blood [lactate] began to increase in a non-linear fashion.12 13 The threshold was estimated by plotting [lactate] against GXT stage. These graphs were visually inspected to determine the lines of best fit by the two evaluators. The equations for each line were set equal to one another and solved for the point of intersection (Figure 1). The values from each evaluator were averaged. 14* We have also revised our results accordingly on page 8 as follows: *there was excellent agreement between estimates of the lactate threshold based on lactate values from the hand-held lactate meter compared to those from the bench top analyzer (r = 0.97). Moreover, there was neither a proportional bias (95% CI for slope: 0.910 to 1.098), nor a fixed bias (95% CI for y-intercept: -0.396 to 0.325) in*

	estimates of the lactate threshold from the hand-held analyzer. The high measurement precision that authors searched for the Lactate Plus lactate meter should also be present in the methodologies used for Lactate Threshold assessment. What is the interest of using a valid and reliable apparatus if the methodology for anaerobic threshold assessment used is subjective (visual inspection)? Gaskill et al (2001) approach helps to achieve higher scientific standards, but mathematical approaches are described in the literature. This fact should not be ignored, so a brief reference in the discussion and limitations sections is welcome. In addition, authors made reference to the equations for the lines of best fit but they were not displayed in Fig 1. 26. The use of fixed blood lactate concentrations of 2.5 and 4.0 mM/l should be justified. Why not 3.5 mM/l, as proposed by Heck et al (Int J Sports Med 1985; 6: 117-130) for lactate threshold, or 8.0 mM/l that is considered a good indicator of aerobic power? The reviewer's point is well taken. Many investigators use several different absolute lactate values to quantify blood lactate concentration. We have added references to support our use of 2.5 and 4.0 mM.l-1 on page xx as follows: *These equations were also used to calculate the stage that corresponded to an absolute blood [lactate] of 2.5 and 4.0 mM.l-1.14 15* This remark was not answered properly. Please justify the use of those fixed blood lactate concentrations (2.5 and 4.0 mM/l, and not, for instance, 3.5 mM/l), and, in our opinion, more important, why an individual value was not determined (as individualized methodologies are available since long time: cf. Stegmann et al. Int J Sports Med 1981; 2: 160 – 165). In fact, comparing the assessment of Lactate Threshold using fixed vs individualized methodologies can evidence differences higher than 1 mM/l that are very important for exercise prescription. 27. Fig 1: if this is an example of a subject please clearly state it. Moreover, if is important to check if the number of points for the YSE and Lactate Plus are correct (6 and 8, respectively). We have revised the figure legend on page 16 to indicate these data are from a study participant as follows: *Figure 1. Determination of the Lactate threshold by visual inspection. Shown are data from a representative study participant and the lines of best fit that were determined independently for data from the YSI 2300 lactate analyzer and the Lactate Plus lactate meter.* We were not able to collect any blood after stage 4 and could not get a blood sample with the hand-held analyzer after stage 7. Thus, the YSI data set contains 9 data points and the Lactate Plus data set contains 8 points. Values for rest and stages 1 and 2 are nearly identical and are difficult to distinguish. Readers will need the above-referred explanation to better understand Fig 1. Please include it (using a more synthetic text). 29. "However, differences of almost 1.0mM can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy". This topic should be better developed since it is important to evidence why differences of ~1.0mM/l are so important for training characterization.

	We appreciate the point made by Dr. Fernandes. We have developed our point more fully in the Discussion on page 10 as follows: *However, differences of almost 1.0 mM.l-1 can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy. Weltman et al. reported that women who trained at an intensity corresponding to about 2.5 mM.l-1 showed greater improvement in blood lactate parameters, but less of an improvement in VO2max than did women training at their lactate threshold. 15 If true,then an error in the measurement of blood lactate concentration could lead to suboptimal improvements in either lactate parameters or VO2max.* Good job, but please observe the comment #26. 30. Although not being the main focus of the current research, it seems important to give the mean (SD) values for blood lactate concentrations corresponding to lactate threshold. As referred in the text, this parameter is of fundamental importance for endurance athletes; so, it should be presented (and discussed). The purpose of estimating the lactate threshold was to determine if the proportional bias we anticipated seeing was large enough to affect the estimation of the lactate threshold or other lactate parameters found in the literature. Thus, it seems to us that the mean value and variability of lactate thresholds within our study sample irrelevant to the aims of our study and interpretation of our data. If we were trying to draw some conclusion about the "eliteness" of our study sample it would certainly make sense, but that is not the case here. It is not irrelevant, as, if the values are in accordance with the literature, it will justify the (subjective) methodology used for lactate threshold assessment. Moreover, in Fig 1, it can be seen that, for one subject, blood lactate concentrations corresponding to anaerobic threshold are between 3-4 mM/L and that it occurs at the 5th or 6th step of the GXT protocol. So, mean (SD) values are welcome.

REVIEWER	Cosme Franklim Buzzachera, Ph.D. Associate Professor, North University of Parana, Londrina, Brazil
REVIEW RETURNED	22-Dec-2012

GENERAL COMMENTS	The authors are to be commended for their consideration and incorporation of the reviewer comments. In doing so, an already strong manuscript was improved.

VERSION 2 – AUTHOR RESPONSE

Authors have done a good job in following the reviewers' suggestions. The manuscript is stronger now. Nevertheless, there are some points that still should be addressed. Please observe the under referred comments. The points that were adequately followed by the authors were deleted to facilitate the reviewing process.

Response: The authors would like to thank Dr. Fernandes for his continued time and effort in helping us improve our manuscript. While we disagree with some of Dr. Fernandes' opinions, we have none-the-less made a good faith effort to incorporate his suggestions where possible. We have detailed our

responses below, providing the exact changes to the manuscript when possible, where each change can be found within the manuscript, or our rationale as to why we have not revised the manuscript as suggested by Dr. Fernandes.

1. Title could be briefer.

Response: We appreciate Dr. Fernandes' sentiments, and would like to have a more concise title as well. However, we have not been able to devise a title of less than 17 words that is both adequately descriptive of the study and meets the journal's requirement to include the study design in the title. We hope that Dr. Fernades will note that our title falls well short of the 50-word limit set by the publishers of BMJOpen.

This reviewer is perfectly aware of the BMJOpen guidelines. Please note that the previous suggestions were not based on personal sense or believes but on scientific writing rules. One of those states that when writing a scientific paper, the language should be completely clear and concise. In this case, there are some words in the title that do not give any additional information to the readers, as it is evident that this is a "study" and that the Lactate Plus is a "hand-held lactate meter".

Response: We have revised the title as follows: A method-comparison study regarding the validity and reliability of the Lactate Plus© analyzer

2. "The Lactate Plus portable lactate meter provides valid and reliable measurements of blood lactate concentration". Is this new? Tanner et al (2010) did not report it before?

Response: Tanner did conclude that the lactate Plus meter "displayed good reliability and accuracy…" However, Tanner's conclusions rely on a questionable analytic approach. Moreover, there clearly appears to be systematic measurement error that was not examined. If one looks at their Figure 4 (shown below) it appears as though a proportional bias exists. This is more evident in Tanner's figure 5 (also shown below). Our approach does not suffer from the assumptions inherent in Tanners analytical approach. Our use of least-products regression allows the reader to assess the accuracy and reliability based on three independent parameters: 1) correlation coefficient; 2) the degree of proportional bias; and 3) the degree of fixed bias. While we come to the same conclusion as Tanner et al, our conclusions are based on a firm analytical approach. Moreover, our approach indicates the meter is 93% more accurate than reported by Tanner et al. We discuss these points in the Discussion on page 10 as follows: Though not specifically assessed, it does appear that Tanner's reported difference between the hand held and reference analyzer is significantly influenced by a proportional bias (Ref 8,Figures 4 and 5). The fact that our data shows little proportional bias (Figure 3) may account for the greater agreement between analyzers that we observed. It is possible that if Tanner had been able to independently determine the proportional and fixed bias, their analysis may have revealed a small bias similar to ours. Differences in reference instruments would not likely explain the greater measurement error reported by Tanner, given that their instrument undergoes a 3- point calibration and 2-point calibration check every few hours, similar to our reference instrument. Moreover, given that Tanner's figures showed a strong proportional bias, as does most other validity data from various hand-held analyzers, our study took the next obvious step, and tested whether this proportional bias was enough to affect the detection of the lactate threshold.

The inclusion of these sentences in the Discussion section better justifies the pertinence of the study.

Response: Thank you

10. “This level of disagreement could be explained by the presence of systematic error, which has gone unexamined in previous studies”. As Baldari et al (ref#2) has examined systematic error, authors are advised to rewrite this sentence. In addition, authors also refer two studies (refs #9 and 10) that studied this topic.

Response: We respectfully disagree. Perhaps the authors and Dr. Fernandes are using the term “systematic error” differently. We explicitly define our use of the term in the introduction based on the definition of Ludbrook (refs 10 and11). We also clearly describe the biases produced by systematic measurement error, which previous studies have not examined. For example, it is clear to us that the data from Baldari et al. displays a proportional bias as shown in their Figure 2, shown near the top of the next page. Moreover, in Figure 3 Baldari reports regression slopes between 0.938 and 1.105, yet does not report whether these slope are significantly different from 1.0 (indicative of a proportional bias). Thus, Baldari did not look for evidence of systematic measurement error in their data. Furthermore, references 9 and 10 (now refs 10 and 11), to which Dr. Fernandes refers do not examine systematic measurement error in hand-held lactate meters, but are papers describing the advantages of least products regression over the least squares regression approach used by Baldari and most other authors that have performed validation studies on these hand-held meters.

The question is not this reviewer interpretation of the term “systematic error” but the existing literature on that topic that uses precisely that expression. So, aiming for a clearer text, the above explanation should be given (in a synthetic way) in the manuscript (eventually in the Discussion section).

Response: We have revised the Introduction to more explicitly state our case as follows: This level of disagreement could be explained by the presence of systematic measurement error. Systematic measurement error can result in a proportional bias, where one instrument produces values that are different from those of another instrument by an amount that is proportional to the level of the measured variable, and/or a fixed bias, where one instrument gives values that are different from those of another instrument by a constant amount.10 11 Thus, similar mean values between lactate analyzers could occur while the portable analyzer produces low values at lower [lactate] and high values at higher [lactate] or vice-versa. Previous studies have primarily relied on Bland-Altman analysis to determine the presence of any fixed bias. However, this approach does not allow the independent determination of bias, and thus has limited utility in assessing the presence of systematic measurement error. Therefore, while most data appear to show a substantial proportional and/or fixed bias the presence and degree of bias in portable lactate analyzers remains unresolved.1 3-8 We believe this revision along with a brief reiteration of these points in the discussion provide a clear and concise explanation of this issue.

15. Lactate analysers are, as referred by the authors, a very important instrument to help in training control and prescription of endurance athletes. Nevertheless, the subjects used in the current study do not seem representative of the high trained athletes. This fact could lower the overall quality of the paper.

Response: We disagree that the training status of the study participants has any relevance to this paper. The aim of this study was to assess the accuracy and reliability of the Lactate Plus analyzer. It is unclear why the device would accurately measure lactate concentration in one population and not in another.

Two questions arise here. The 1st is related to the authors´ writing style: if the Introduction starts with

reference to the fact that “blood lactate accumulation is a common measure in the physiological assessment of endurance athletes”, naturally the readers will maintain this target population in mind, and will find confusing that no athletes were used during the experimental procedures. Reference to “training intensity or efficacy” when using blood lactate concentration values is also presented in the Discussion section. The 2nd question is related to the usefulness of the tests conducted using the Lactate Plus. If it will be mainly used for assessing lactate threshold (as often referred by the authors along the text) it should be for training control and evaluation purposes. So, as the subjects involved in training process are athletes, the population used in the current study is very relevant. This should be assumed as a study limitation.

Response: Comments regarding writing style are difficult for us to process and implement given the vastly different opinions of the two reviewers. While Dr. Fernandes finds the style confusing or verbose, Dr. Buzzachera stated that, “The authors are to be commended for a well-written manuscript.” None-the-less, we have taken Dr. Fernandes’ opinions into consideration and have revised a few sections.
While we continue to disagree with Dr. Fernandes’ opinion regarding our use of non-elite athletes, we have added a section in the Discussion on page 13 to help the reader consider this issue in the context of previous validation studies as follows: While some studies have used blood collected from trained athletes to compare portable lactate analyzers to bench top models, 5 6 8 10 several do not. 3-5 7 9 This seems quite appropriate given that the importance of accurate lactate measurement extends well beyond the athletic field. Our subjects were healthy and physically active, but not highly trained. This is unlikely to account for any difference between previous studies and ours given that we can find no reason to speculate that either lactate analyzer would more accurately measure [lactate] in one population compared to another.

22. Please explain it were not used fixed protocol increments. Was this protocol previously described in the literature?

Response: This GXT protocol has not been previously described in the literature. It is unclear to the authors how our protocol would negatively affect our ability to assess the accuracy and reliability of the hand-held analyzer, or model changes in blood lactate concentration.

The assessment of lactate threshold and other physiologic parameters are strongly affected by the used protocol (there are several examples of lactate threshold assessment protocols in the literature). Thus, the GXT protocol should be better described (e.g. why it were not used fixed protocol increments in the first steps) and discussed with similar (already published) approaches.

Response: We have added a statement in the Discussion (pg. 13) reflecting the fact that our choice of protocol likely affected the estimation of the LT as follows: Similarly, the choice of graded exercise protocol can affect lactate threshold determination. 29 Thus, our use of a personalized, discontinuous GXT likely produced LT values different from some other protocols. However, this would have no affect on our ability to accomplish the aims of our study, specifically to compare estimates of LT between lactate measurements produced by the portable and reference analyzers.

23. “…Bland-Altman plots were constructed to allow the reader to …” Authors choose their statistical procedures based on scientific principles or in the readers opinion? Please rewrite.

Response: We agree that statistical analyses should be chosen based on the experimental question

or hypothesis being tested and statistical principles. However, we also believe that an important aspect of writing a scientific paper is to inform the readers. This includes helping the readers understand our findings within the context of previous work. Therefore, we chose to construct a Bland-Altman plot because this has become commonplace in methodological studies. (see ref 1-6,8) So as a service to our readers, we provide a common point of comparison between our data and those previously published.

Off course it is important to inform the readers. What should be avoided is to refer it (it is implicit). Please rewrite.

Response: It is unclear to us what Dr. Fernandes means by “refer it”. Do you mean infer? Are you asking us to state more explicitly why we constructed a Bland-Altman plot? We already explicitly state in our Methods section on page 7 that the “Bland-Altman plot was constructed to allow the reader to more directly compare our data to that of previous studies since this is the approach typically used.”

24. The 1st paragraph of the Data Analysis section is too descriptive. In our opinion, it should be briefer and some references should be added.

Response: Based on the analytic approaches used in previous validation studies, it is reasonable to assume that a thorough explanation of our approach is warranted. We have added references to this section on page 7 as follows: Validity was determined from the correlation coefficient in combination with the presence and degree of bias. The degree of fixed bias was determined from the y-intercept 95% confidence intervals. If the confidence interval for the intercept includes the value of zero, then there is no fixed bias. Proportional bias was determined from the 95% confidence interval for the slope. If the confidence interval for the slope includes the value of 1.0, then there is no proportional bias. Ordinary least products regression gives different slopes and y-intercepts than does least squares regression because error is assumed in both hand-held and bench top analyzers. 10 11

The writing style should be perfected, per example by merging consecutive sentences: “Validity was determined from the correlation coefficient in combination with the presence and degree of bias (determined from the yintercept 95% confidence intervals)”. This should be done for all the manuscript.

Response: Thank you for prompting us to take another look at our manuscript. We have taken Dr. Fernandes’ opinion into consideration and revised several sections.

25. As it is well described that after lactate threshold intensity of blood lactate concentrations assumes an exponential increase, we wonder if the use of 2 linear regressions in the best way to assess lactate threshold. If authors want to go deep in this topic, the can consult a study of our group (Fernandes RJ et al. Individual Anaerobic Threshold in Swimming, Int J Sports Med 2011; 32: 940-946).

Response: This is the one common concern shared by the reviewers, and we agree this is an issue that needs to be addressed. We chose to follow the procedures outlined by Gaskill et al (Med Sci Sports Exerc 2001; 33(11):1841-48) as suggested by Dr.Buzzachera. This has slightly reduced the correlation coefficient and changed the parameters of the regression line. We have clarified our approach in the Methods section on page 7 as follows: Lactate threshold was defined as the point at which blood [lactate] began to increase in a non-linear fashion.12 13 The threshold was estimated by plotting [lactate] against GXT stage. These graphs were visually inspected to determine the lines of best fit by the two evaluators. The equations for each line were set equal to one another and solved

for the point of intersection (Figure 1). The values from each evaluator were averaged. 14 We have also revised our results accordingly on page 8 as follows: there was excellent agreement between estimates of the lactate threshold based on lactate values from the hand-held lactate meter compared to those from the bench top analyzer (r = 0.97). Moreover, there was neither a proportional bias (95% CI for slope: 0.910 to 1.098), nor a fixed bias (95% CI for y-intercept: -0.396 to 0.325) in estimates of the lactate threshold from the hand-held analyzer.

The high measurement precision that authors searched for the Lactate Plus lactate meter should also be present in the methodologies used for Lactate Threshold assessment. What is the interest of using a valid and reliable apparatus if the methodology for anaerobic threshold assessment used is subjective (visual inspection)? Gaskill et al (2001) approach helps to achieve higher scientific standards, but mathematical approaches are described in the literature. This fact should not be ignored, so a brief reference in the discussion and limitations sections is welcome. In addition, authors made reference to the equations for the lines of best fit but
they were not displayed in Fig 1.

Response: Your argument regarding precision in our explanation is well taken. Thank you. In response we have added details regarding our analytical approach on pages 7-8 as follows: Lactate threshold was defined as the point at which blood [lactate] began to increase in a non-linear fashion.14 15 The threshold was estimated by plotting [lactate] against GXT stage. These graphs were visually inspected to determine the lines of best fit by the two evaluators. The following guidelines were used to help guide the evaluators: 1) at least 3 data points were included in each line, 2) both lines contained unique data points, and 3) lines were chosen that produced the highest R2 with the smallest confidence intervals. Once the lines were chosen the equations for each line were set equal to one another and solved for the point of intersection (Figure 1). The values from each evaluator were averaged.16
We have also revised the Discussion on pages 11-12 to explore the potential shortcomings of our approach versus methods purported to be more objective for assessing the LT as follows: Determination of the LT by visual inspection has come under scrutiny. 21 22 To reduce subjectivity our approach to visual inspection is guided by several principles similar to those used by others. 16 23 Several methods of assessing the LT have been proposed that purport to be more objective. 14 16 24 However, many of these methods are known to be significantly affected by data outliers and/or missing data. 25 26 Therefore, the choice of any analytical approach has a subjective component. While our approach likely produces LT values that are different from other approaches, it produced values consistent with other studies that employed similar approaches to LT estimation. 18 23 When one considers the strong correlation and small biases in our data, it seems likely the LT estimates would be strongly correlated regardless of the analytical approach chosen.

26. The use of fixed blood lactate concentrations of 2.5 and 4.0 mM/l should be justified. Why not 3.5 mM/l, as proposed by Heck et al (Int J Sports Med 1985; 6: 117-130) for lactate threshold, or 8.0 mM/l that is considered a good indicator of aerobic power?

Response: The reviewer's point is well taken. Many investigators use several different absolute lactate values to quantify blood lactate concentration. We have added references to support our use of 2.5 and 4.0 mM.l-1 on page xx as follows: These equations were also used to calculate the stage that corresponded to an absolute blood [lactate] of 2.5 and 4.0 mM.l-1.14 15

This remark was not answered properly. Please justify the use of those fixed blood lactate concentrations (2.5 and 4.0 mM/l, and not, for instance, 3.5 mM/l), and, in our opinion, more important, why an individual value was not determined (as individualized methodologies are available since long time: cf. Stegmann et al. Int J

Sports Med 1981; 2: 160 – 165). In fact, comparing the assessment of Lactate Threshold using fixed vs individualized methodologies can evidence differences higher than 1 mM/l that are very important for exercise prescription.

Response: Regarding point 1, we have revised the Discussion on page 11 to read as follows: These lactate concentrations were chosen because they have both sport and clinical significance. 1 2 19 20 The strong correlation coefficient and small biases suggest that the Lactate Plus analyzer can be used to accurately determine exercise intensities based on any blood lactate parameter.
Regarding your second point, what Stegmann and colleagues suggest at the end of their Discussion is that a fixed blood lactate concentration may be a poor indicator of endurance capacity. This point would be pertinent if the aims of our study were to develop or compare methodologies for determine lactate threshold. Furthermore, we estimate both individual (LT) as well as fixed (2.5 and 4.0 mmol/l) parameters of lactate kinetics.

27. Fig 1: if this is an example of a subject please clearly state it. Moreover, if is important to check if the number of points for the YSE and Lactate Plus are correct (6 and 8, respectively).

Response: We have revised the figure legend on page 16 to indicate these data are from a study participant as follows: Figure 1. Determination of the Lactate threshold by visual inspection. Shown are data from a representative study participant and the lines of best fit that were determined independently for data from the YSI 2300 lactate analyzer and the Lactate Plus lactate meter. We were not able to collect any blood after stage 4 and could not get a blood sample with the hand-held analyzer after stage 7. Thus, the YSI data set contains 9 data points and the Lactate Plus data set contains 8 points. Values for rest and stages 1 and 2 are nearly identical and are difficult to distinguish.

Readers will need the above-referred explanation to better understand Fig 1.
Please include it (using a more synthetic text).

Response: This information has been added to the figure legend as follows: Determination of the Lactate threshold by visual inspection. Shown are data from a representative study participant and the lines of best fit that were determined independently for data from the YSI 2300 lactate analyzer and the Lactate Plus lactate analyzer. Blood samples could not be collected between stages 4 and 5. The Lactate Plus analyzer returned error message between stages 6 and 7.

29. “However, differences of almost 1.0mM can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy”. This topic should be better developed since it is important to evidence why differences of ~1.0mM/l are so important for training characterization.

Response: We appreciate the point made by Dr. Fernandes. We have developed our point more fully in the Discussion on page 10 as follows: However, differences of almost 1.0 mM.l-1 can significantly impact the use of absolute [lactate] to characterize training intensity or efficacy. Weltman et al. reported that women who trained at an intensity corresponding to about 2.5 mM.l-1 showed greater improvement in blood lactate parameters, but less of an improvement in VO2max than did women training at their lactate threshold. 15 If true,then an error in the measurement of blood lactate concentration could lead to suboptimal improvements in either lactate parameters or VO2max.

Good job, but please observe the comment #26.

Response: Thank you

30. Although not being the main focus of the current research, it seems important to give the mean (SD) values for blood lactate concentrations corresponding to lactate threshold. As referred in the text, this parameter is of fundamental importance for endurance athletes; so, it should be presented (and discussed).

Response: The purpose of estimating the lactate threshold was to determine if the proportional bias we anticipated seeing was large enough to affect the estimation of the lactate threshold or other lactate parameters found in the literature. Thus, it seems to us that the mean value and variability of lactate thresholds within our study sample irrelevant to the aims of our study and interpretation of our data. If we were trying to draw some conclusion about the “eliteness” of our study sample it would certainly make sense, but that is not the case here.

It is not irrelevant, as, if the values are in accordance with the literature, it will justify the (subjective) methodology used for lactate threshold assessment. Moreover, in Fig 1, it can be seen that, for one subject, blood lactate concentrations corresponding to anaerobic threshold are between 3-4 mM/L and that it occurs at the 5th or 6th step of the GXT protocol. So, mean (SD) values are welcome.

Response: We have added the information in the Results section on page 9 as follows: Regardless of blood sampling approach there was excellent agreement between estimates of the LT based on lactate values from the portable analyzer compared to those from the bench top analyzer ($r = 0.97$). Moreover, there was neither a proportional bias (95% CI for slope: 0.910 to 1.098), nor a fixed bias (95% CI for y-intercept: -0.396 to 0.325) in estimates of the lactate threshold from the portable analyzer. Given the lack of bias it is not surprising there was no difference between blood [La] at the LT (2.88NOVA±0.53 vs. 3.15YSI±0.46 mM.l-1; $p=0.32$). In addition the stages corresponding to absolute blood lactate values of 2.5 mM.l-1 (2.99NOVA vs. 2.92YSI) and 4.0 mM.l-1 (4.64NOVA vs. 4.61YSI) were not different between portable and bench top values ($p = 0.86$ for both).
We have also added references in the Discussion on page 12 as follows: While our approach likely produces LT values that are different from other approaches, it produced values consistent with other studies that employed similar approaches to LT estimation. 18 23